

Article

Analysis of Soil Carbon Stock Dynamics by Machine Learning—Polish Case Study

Artur Łopatka ^{1,*} , Grzegorz Siebielec ¹ , Radosław Kaczyński ¹ and Tomasz Stuczyński ²

¹ Department of Soil Science Erosion and Land Protection, Institute of Soil Science and Plant Cultivation—State Research Institute, Czarzoryskich 8, 24-100 Pulawy, Poland; gs@iung.pulawy.pl (G.S.); rkaczynski85@gmail.com (R.K.)

² Faculty of Science and Health, The John Paul II Catholic University of Lublin, Konstantynów 1 H, 20-708 Lublin, Poland; tomasz.stuczynski@kul.pl

* Correspondence: artur@iung.pulawy.pl

Abstract: A simplified differential equation for the dynamics of soil organic carbon (SOC) that describes the rate of SOC change ($dSOC/dt$) was constructed using the LASSO regression—a regularized linear regression machine learning method. This method selects the best predefined explanatory variables and empirically evaluates the relevant parameters of the equation. The result, converted into a formula for the long-term equilibrium level of soil carbon, indicates the existence of carbon sequestration potential in the studied regions of Poland. In particular, the model predicts high SOC content in regions with a high Topographic Wetness Index (TWI), such as river valleys or areas with high cattle density, as expected.

Keywords: soil organic carbon; livestock; erosion; TWI; Aridity Index; machine learning; differential equation; LASSO



Citation: Łopatka, A.; Siebielec, G.; Kaczyński, R.; Stuczyński, T. Analysis of Soil Carbon Stock Dynamics by Machine Learning—Polish Case Study. *Land* **2023**, *12*, 1587. <https://doi.org/10.3390/land12081587>

Academic Editors: Cole D. Gross, Zhengfeng An and Scott X. Chang

Received: 10 July 2023

Revised: 8 August 2023

Accepted: 10 August 2023

Published: 11 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Agronomists aiming to improve the agricultural quality of soils have long been interested in the potential to increase soil carbon resources. Climatologists study both the threats related to the degradation of soil organic matter and the opportunities related to the potential for sequestration of excess carbon dioxide from the atmosphere in soils [1–3].

Most studies on changes in organic carbon resources in soils use mechanistic models that use systems of differential equations [4]. The first-order kinetics equations describe the rate of changes in the stocks of several pools of organic matter as a balance of two processes: accumulation with a positive contribution and decomposition through mineralization with a negative contribution. The sources of exogenous carbon are post-harvest plant residues and organic fertilizers that feed the accumulation of the first pool of organic matter that decomposes quickly. Some of the products of this decomposition feed into the process of accumulation of the next slower decomposing pool. Since the rate of mineralization increases with the increase in carbon stocks in the soil, the equations describing the dynamics predict the existence of a certain level of carbon stocks at which the rate of mineralization and accumulation is equal. Many mechanistic models have been developed [4–7]. They differ in the number of separated organic matter pools and the role of microbial activity [8,9]. They also differ in relations describing the influence of soil temperature and soil moisture on the rate of mineralization [4] or the impact of saturation capacity related to clay fraction and mineralogy of soil [10,11]. A significant part of the new research focuses on models analyzing in detail soil microbial processes described by the Michaelis-Menten enzymatic kinetics equation [9,12]. However, it is currently unclear to what extent these more complex models offer better predictions of SOC stock changes [8,13–16].

Mechanistic models, trying to reflect the physical and chemical processes taking place in nature, are complex for obvious reasons and require the determination of many empirical

parameters depending on the farmer's decisions. These include long-term time series data on the type of crops, yields, amount of post-harvest residues left on fields, level of organic fertilization, method of soil tillage, etc. This data is easily collected for single-field or farm-scale simulations. The typical range of agricultural input data required by mechanistic simulation models is shown in the example of the three most frequently cited models in the Web of Science database [17]: RothC, CENTURY, and DNDC. In the RothC model [18–20], these are monthly soil cover (bare or vegetated), monthly residue C, and farmland manure C inputs. In the CENTURY model [19,21,22], these are crop type, management events (sowing, tillage, harvesting), and plant C and N content. In the DNDC model [23,24], these are the tillage method, N fertilization rates, planting and harvest dates, amount of post-harvest residues, and maximum yield. On the scale of macro analyses, detailed data on C loads entering the field are difficult to access and, therefore, must be estimated. An example of such estimation is the use of data on livestock density and yields in large administrative units (for Poland, these are voivodeships classified in the EU as NUTS2 units). In particular, there are no models for the distribution of organic fertilizers in the vicinity of farms with large livestock production or the share of post-harvest residues left on the fields. For example, in the globally applied RothC model [20], data on crop residues were estimated using yield data reported by FAO at the country level. The model assumed constant shares of above-ground crop residues and underground residues in relation to above-ground yield. The amount of carbon entering the fields in livestock manure was determined based on stocking density layers for cattle, pigs, and poultry for the world [25] in a raster with a resolution of 3 arcminutes (5 by 5 km at the equator). Although the data on stocking density collected in administrative units were disaggregated, in Poland, the predicted stocking density shown on this map was almost homogeneous within voivodeships. As a result, abrupt changes are visible at the borders of voivodeships. In some areas, therefore, the actual resolution of stocking density layers is determined mainly by the size of the administrative units in which the data is reported. Since most models are not validated against an independent time series of SOC measurements [7], it is difficult to find a model that can assess the prediction errors associated with the use of macro-estimated data. However, it has been shown that the high uncertainty of the modeling results of SOC changes on agricultural land is associated with the choice of the crop carbon inputs estimation method [26]. It has also been observed that the main source of errors in global SOC models for natural ecosystems is errors in the estimation of the net primary productivity [27], which determines the amount of C input.

Therefore, despite good recognition and the possibility of reflecting soil organic matter processes in the models, the level of carbon resources in the soil is strongly affected by parameters depending on the farmer's decisions. They are estimated with high uncertainty, particularly in analyses of large areas. However, it is known that farmers adapt both sowing structure, level of fertilization, method of soil cultivation to soil texture, stocking density, etc. Since changes in soil carbon stocks are slow, the information provided to the model by a time series of annual cropping or organic fertilization data may be replaceable by the information contained in causally related variables such as average livestock density or soil texture. The transition to these new less numerous variables in the mechanistic model is difficult due to the many original variables that depend on them. Therefore, it may be more practical to replace whole parts of the mechanistic model with empirically estimated functions.

In connection with these observations, we hypothesize that it may be useful to simplify the structure of mechanistic soil organic carbon models to a form in which their parts can be constructed from variables that are easily accessible and convenient for estimation and interpretation. In the Materials and Methods chapter, we describe (i) our attempt to simplify the equations describing the mechanistic model, (ii) the method of determining the empirical parts of the model using machine learning methods, (iii) the formula for the equilibrium level of carbon in soil resulting from the model, and (iv) the measurement data applied to determine the empirical parts of the model. In the next chapters, Results and Discussion,

we discuss (i) the structure of the model, (ii) the structure of the equilibrium level of carbon in the soil, (iii) the cross-validation model results, (iv) the comparison of predictions of the obtained model with findings of previous research, and (v) the implications of model findings for practice. The identified gap in the existing mechanistic models, related mainly to errors in the assessment of C inputs, is attempted to be filled using a simple first-order kinetics equation. Unlike in existing models based on such equations, explanatory variables are selected automatically and the parameters determining the influence of these variables on the model are estimated by regression methods similarly to empirical models.

2. Materials and Methods

2.1. Study Area

We analyzed 228 sampling points (Figure 1) located in central Europe (51.6–54.0° N, 14.8–23.8° E) grouped into four administrative regions of Polish voivodeship (NUTS2 level): *a*—Podlaskie, *b*—Lubelskie, *c*—Zachodniopomorskie and *d*—Dolnoslaskie. The terrain is lowland (25–339 m a.s.l.), mostly flat, and slightly undulating, except for the southern measurement points in region *b*, which are in strongly undulating terrain. According to the Köppen-Geiger classification, the climate is humid continental [28] with the ratio of average annual precipitation to potential evapotranspiration AI (Aridity Index) ranging from 0.76–1.18 and with a gradient of average temperature towards the southwest (region *d*) to the northeast (region *a*). The range of the sum of Growing Degree Days (GDD), i.e., effective temperatures above 5 °C, ranges from 1647 degree-days in region *a* to 2044 degree-days in region *d*, which in agricultural practice translates into a delay in the beginning of the growing season in region *a* relative to region *d* by about 2 weeks. The post-glacial soils are strongly differentiated in terms of grain size, with a predominance of sandy soils (loamy sands and sandy loams according to the USDA classification) with the share of floatable fraction (>0.02 mm) ranging from 3.5 to 66% with an average of 24.2%. The largest share of sandy soils is found in regions *a*, *c*, and the north of region *b*, while the south of region *b* and most of area *d* are composed of silty loams.

In region *d*, which is the warmest and has loess soil with high agricultural suitability, intensive agriculture prevails. This is evident from the crop classification map for 2018 [29], where the total share of demanding crops such as wheat, barley, rapeseed, and sugar beet in points of region *d* was about 77%. In contrast, in the coldest and sandy region *a*, this share was only about 38%. The most unfavorable agricultural conditions for intensive plant cultivation in region *a* contributed to the fact that the changes in Polish agriculture that took place after the departure from the centrally planned economy in 1989 favored the concentration of cattle breeding in this region, while in other regions, the number of cattle was reduced. In the last 20 years, the average stocking density of cattle in administrative units of the NUTS5 level (PSR 2002, PSR 2020) in which the measurement points of region *a* are located was about 0.76 livestock units LU ha^{−1}, while for region *d*, it was only 0.07 LU ha^{−1}, and for regions *b* and *c*, it was 0.20 and 0.10 LU ha^{−1}, respectively. Cattle breeding is conducive to maintaining grassland and a structure of crops on arable land in which fodder crops have a significant share. This ensures that soil has dense vegetation cover throughout or for most of the year and is well protected against erosion. Water erosion is a particular threat in areas with loess soil that are susceptible to leaching, and in the study area, it is most intense in the south of region *b*.

We expect that a high stocking density, a large share of grassland, and a relatively cool and humid climate should be conducive to the accumulation of organic carbon in the soils of region *a*. However, in region *d*, which is almost completely devoid of cattle and grasslands, and in the loessive south part of region *b*, which favors water erosion, the accumulation is assumed to be low.

The performed regionalization of research should, therefore, ensure conditions in which the range of SOC changes in a relatively small research area is maximized and, thus, facilitate the correct estimation of the parameters of the tested model.

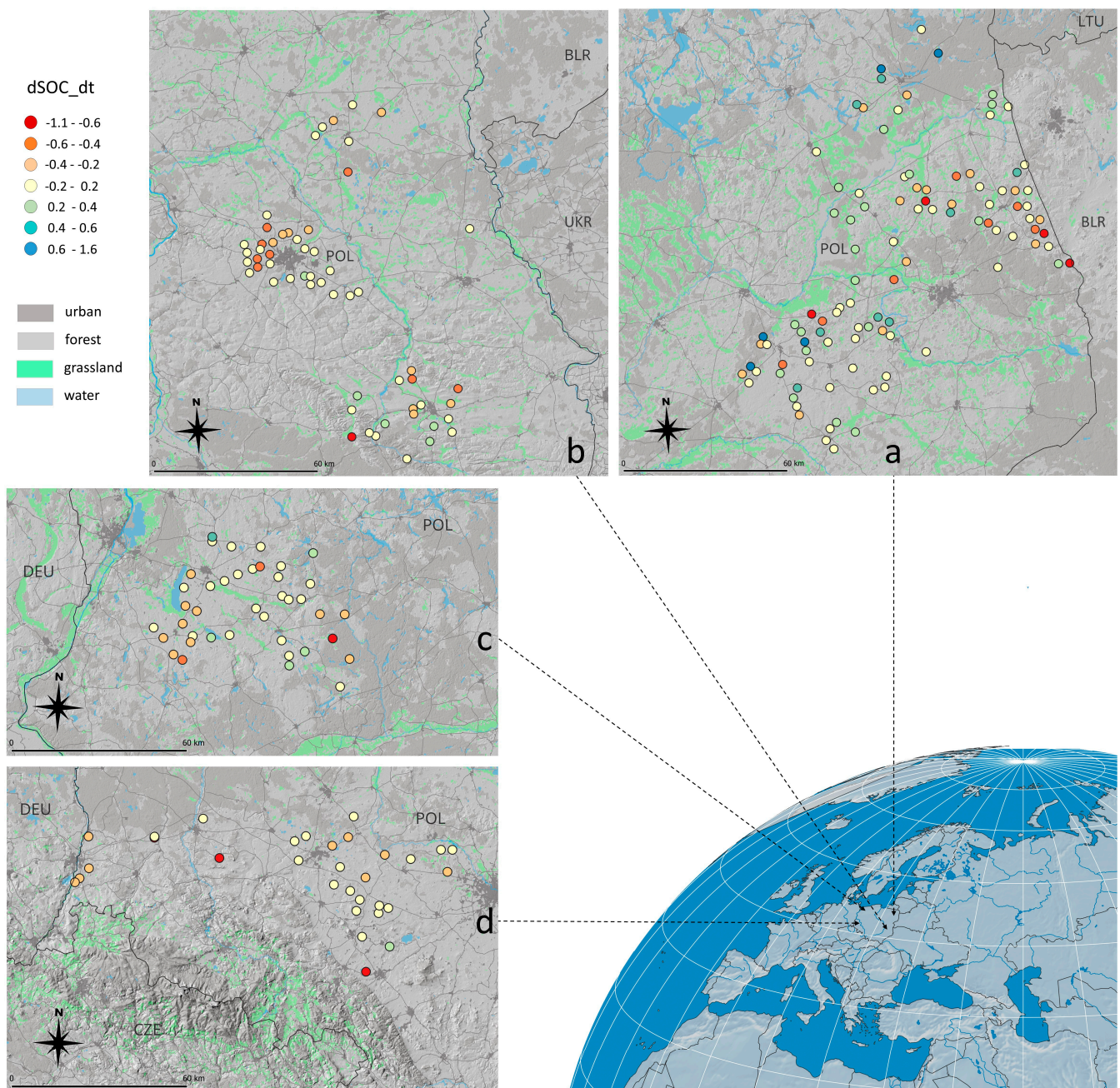


Figure 1. Localizations of measurement points and means of soil organic carbon stock change rate $dSOC_{dt}$ from two periods in regions: a—Podlaskie, b—Lubelskie, c—Zachodniopomorskie and d—Dolnoslaskie.

2.2. Soil Data Preparation

Sampling locations were selected from the two existing databases of SOC measurements in arable soil profiles. The first database contains 11,000 soil profiles collected for the preparation and quality control of the agricultural soil map of Poland in the 1970s and 1980s. The second database was collected mainly in 2003–2004 at selected locations of the first database for the purpose of assessing changes in soil properties. In 2018, the measurements were repeated, obtaining complete data for three periods at 228 points, including exact location and measurements of SOC% content and grain size in at least the upper 30 cm of soil.

All measurements of soil organic carbon level and soil texture were made in samples taken from a depth of 0–30 cm. Determinations of soil organic carbon content SOC% were made using the Tiurin method [30] and determination of granulometric composition by Casagrande's areometric method in modification by Prószyński [31].

In each of the points, two dSOC changes were calculated from three SOC measurements taken in the years 1969–1989, 2002–2004, and 2018. These changes were divided by the time period Δt to give two values of the dSOC_Δt change rate.

The soil carbon level SOC% was converted to its stock SOC (t ha^{-1}) in the 0–30 cm layer using the formula:

$$SOC = d \cdot BD \cdot SOC\% \quad (1)$$

where depth $d = 30$ cm and the bulk density BD (g cm^{-3}) value was calculated according to a pedotransfer formula for soils in Poland [32].

The average stock of soil organic carbon SOC in the analyzed points, located mainly on arable lands, was 43.3 t ha^{-1} (minimum value 24.5, median 40.8, and maximum value 87.4 t ha^{-1}). These values are slightly smaller than average SOC values reported for arable land and grassland soils of Poland [33] on a larger dataset ($N = 50,000$), where minimum, median, mean, and maximum values of stock for the same depth of 0–30 cm were, respectively, 35, 58, 59, and 97 t ha^{-1} . The average rate of change dSOC_Δt in the studied points was negative, i.e., soils lost organic carbon at an average rate of $-0.06 \text{ t ha}^{-1} \text{ year}^{-1}$ (minimum, median and maximum were, respectively, equal to -1.98 , -0.08 , and $3.87 \text{ t ha}^{-1} \text{ year}^{-1}$). The negative but close to zero balance of organic matter observed in the studied points is consistent with current estimates for Poland [34] based on reproduction-degradation coefficients proposed by Eich and Kundler. For the years 2017–2019, these predicted that dSOC_Δt is positive and equal to $0.07 \text{ t ha}^{-1} \text{ year}^{-1}$.

2.3. Analysis Overview

In order to maintain a dynamic description offered by mechanistic models using systems of differential equations and their description of reaching the steady-state level with the equilibrium of mineralization and accumulation, their maximum simplification was made to a description by means of one equation for all pools of soil organic matter.

The simplest such description of soil organic carbon stock changes dSOC/dt ($\text{t ha}^{-1} \text{ year}^{-1}$) is possible with a single pool model of first-order kinetics [17,35,36]:

$$\frac{dSOC}{dt} = I - k \cdot SOC \quad (2)$$

where I ($\text{t ha}^{-1} \text{ year}^{-1}$) is the accumulation rate and k (year^{-1}) is the SOC decomposition coefficient. The above first-order linear differential equation with positive coefficients I and k approximately independent of time t over period Δt considered in the analysis, was discretized by the trapezoidal rule:

$$\frac{dSOC}{dt} \approx \frac{SOC_{t+\Delta t} - SOC_t}{\Delta t}; \quad SOC \approx \frac{SOC_{t+\Delta t} + SOC_t}{2} \quad (3)$$

The coefficients I and k are considered as changing in space and dependent on the set of selected explanatory variables (listed in the next paragraph). Since the coefficients function form is unknown, its approximation was used, which is a linear combination of selected variables in the first and second power. The number of such terms increases rapidly with the number of variables, and they are often correlated. Therefore, we used, one of the machine learning methods proposed for such cases [37]—regularized LASSO regression [38], which results in the selection of the most important variables, model overfitting limitation, and, at the same time, the model interpretability improvement. The parameters of the LASSO regression model were estimated using the glmnet package in R.

The form of model (2), selected by the LASSO machine learning method, can be used to determine the equilibrium SOC level for conditions when SOC changes are negligible:

$$SOC_{eq} = \frac{I}{k} \quad (4)$$

This will be further interpreted in terms of optimizing strategies for increasing carbon sequestration in soils.

A change in SOC may also occur because we did not reach the point of the previous measurement. In this case, not only will the SOC change, but also other soil parameters, including soil particle size distribution. As soil texture is almost constant over time, the expression dFloat_dt with a change in the fraction of soil floatable particles was added to the model Equation (2) as a measure of not hitting the same measurement point. In Polish soils, a floatable fraction is easier to measure (due to a higher share and wider range of variation) than the clay fraction. It should be noted that erosion may also result in the leaching of the floatable fraction and the simultaneous loss of organic carbon in the soil. Therefore, special attention will be required when interpreting the results of the model term with the floatable fraction change rate.

2.4. Explanatory Variables

The choice of explanatory variables was guided by an expert assessment of their importance for the processes shaping the level of soil organic matter and the availability of spatial databases with sufficient resolution. The total set of 8 selected explanatory variables is presented in Table 1.

Table 1. Explanatory variables used in the analysis.

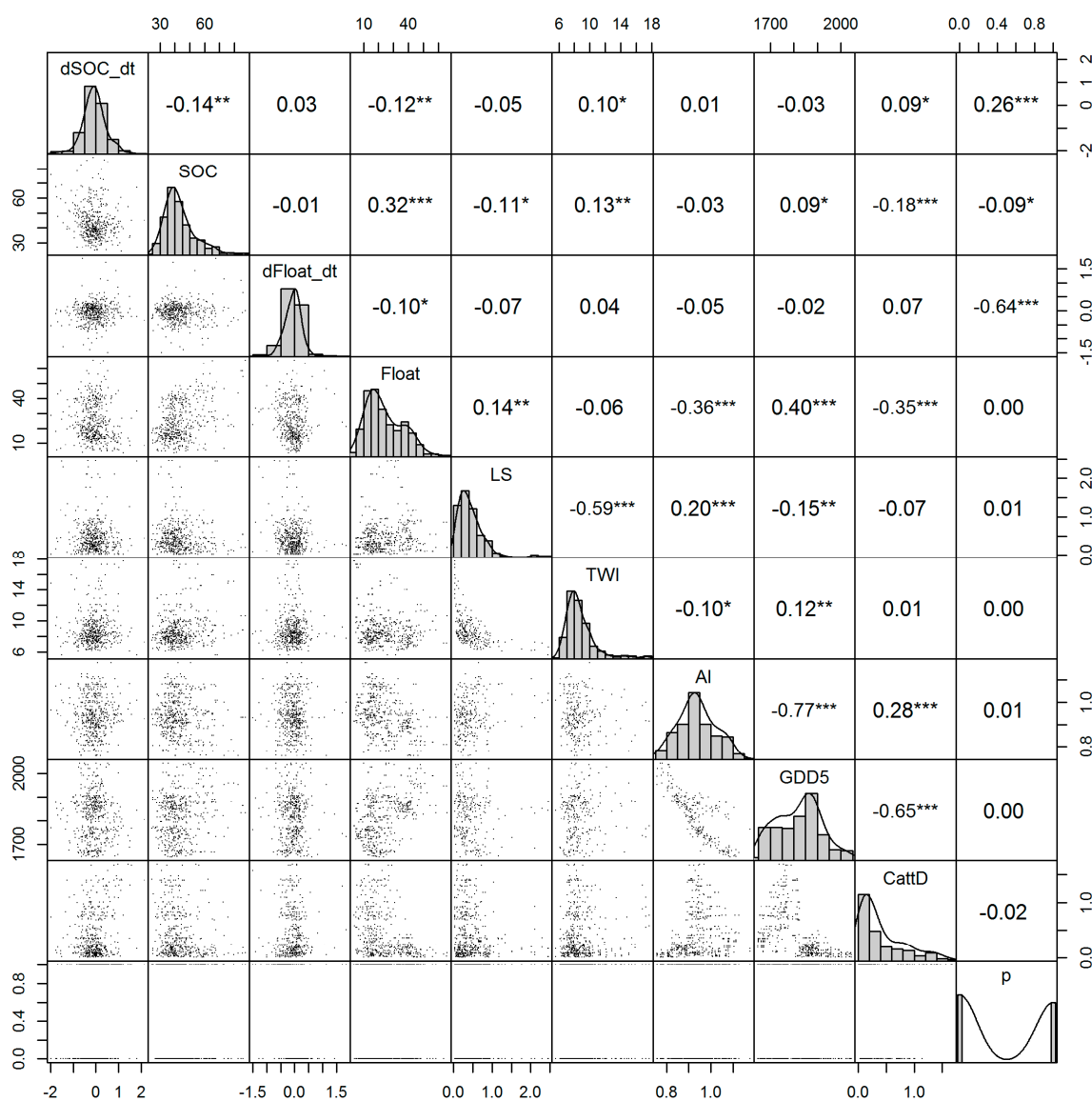
Variable	Definition of the Explanatory Variable and Its Impact on the Level SOC	Range				Data Source
		Minimum	Median	Mean	Maximum	
SOC	Soil organic carbon stock (t ha ⁻¹)	24.5	40.8	43.3	87.4	measured
dFloat_dt	Change rate of floatable soil fraction <0.02 mm (% year ⁻¹)	−1.50	−0.06	−0.06	1.86	measured
Float	Floatable fraction content (%) in the topsoil; determines the soil structure, its retention, and sorption capacity; affects the level of yield	3.5	21.0	24.2	66.0	measured
LS	Length-slope factor in USLE model; based on a digital terrain model with a resolution of 20 m; topographic component of water erosion	0.03	0.36	0.47	2.46	own calculations based on the data EU-DEM [39]
TWI	Topographic Wetness Index (-); calculated on the basis of a digital terrain model with a resolution of 20 m using SagaGIS according to current recommendations [40]; topographic component of soil moisture	5.59	8.26	8.77	17.72	own calculations based on the data EU-DEM [39]
AI	Aridity Index (-) the ratio of annual precipitation and potential evapotranspiration, averaged for the years 1971–2015; the climatic component of soil moisture	0.76	0.93	0.94	1.18	own calculations based on the data IMiGW [41]
GDD5	The sum of effective temperatures (day °C) for the growing season (days with temp > 5 °C) being the sum of daily differences between the average daily temperature and the base temperature equal to 5 °C; averaged for the years 1971–2015	1647	1834	1821	2044	own calculations based on the data IMiGW [41]

Table 1. Cont.

Variable	Definition of the Explanatory Variable and Its Impact on the Level SOC	Range				Data Source
		Minimum	Median	Mean	Maximum	
CattD	Stocking density of cattle (animals ha ^{−1} UAA *); affects the level of doses of organic fertilizers and arable land crop structure	0.01	0.23	0.41	1.68	own calculations based on the data PSR2002 [42] and PSR2020 [43]
p	Dummy variable for differentiating old and new changes of SOC (0—old; 1—new)	0	0	0.47	1	-

* UAA—Utilized Agricultural Area.

Efforts were made to select explanatory variables that are not obviously causally related, which is largely confirmed by the correlation analysis. The Spearman correlations between the variables in the model are usually low (Figure 2), although, in the case of pairs LS-TWI, AI-GDD5, GDD5-CattD, and dFloat_dt-p, they exceed the level of 0.5.

Figure 2. Spearman correlations between model variables (* $p < 0.5$, ** $p < 0.01$, and *** $p < 0.001$).

Despite the existence of a few strongly correlated variables, the use of the LASSO regularized regression method eliminates undesirable correlation effects on the predictive capabilities of the model through its built-in variable selection.

2.5. Validation

The model of changes in soil carbon stock was validated in two ways: (i) using the spatial 4-fold cross-validation method (points were split into 4 parts according to their natural geographical grouping as shown in Figure 1) and (ii) validation on an independent set of data from monitoring of chemistry of arable soils in Poland [44,45] for SOM change in years 2000–2015 where $N = 191$ (after selecting from 216 points only those with complete explanatory data). This set of data was chosen for its open-access policy and high quality of soil sampling, ensured especially by the good homogenization of soil samples consisting of several partial samples taken from an area of approximately 100 m². The accuracy of the model in matching the predicted rate of change in organic carbon stocks $dSOC/dt$ (t ha⁻¹ yr⁻¹) to the observed data was measured by the coefficient of determination R^2 and the root mean square error (RMSE).

3. Results

3.1. Soil Organic Carbon Dynamics Model

As a result of using the regularized LASSO regression method, the following form of the model was obtained:

$$\frac{dSOC}{dt} = (0.408 - 0.139p) \frac{dFloat}{dt} + (-0.104 + 0.293p + 0.009 TWI + 0.083CattD^2) - (0.004 + 0.002AI - 0.002CattD^2) SOC \quad (5)$$

with RMSE = 0.474 and $R^2 = 0.196$. The diagnostics showed (Supplementary Figure S1 and Table S1) the need to remove two outliers, the normal distribution of the model residuals, and the lack of their spatial correlation, which was confirmed by Moran's I test.

Model validation results were (i) RMSE = 0.495 and $R^2 = 0.047$ in the cross-validation method and (ii) RMSE = 2.04 and $R^2 = 0.020$ for the validation on an independent set. The cross-validation results are positive, although they explain a similar small part of the variability of $dSOC_dt$ as a single correlation with a change in grain size of the floatable particle fraction or cattle density (Figure 3).

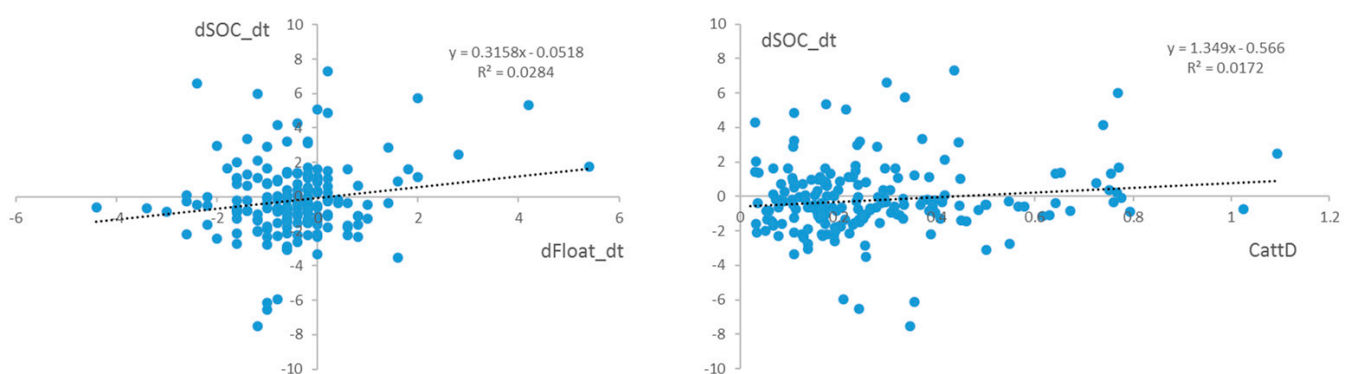


Figure 3. Independent dataset relation between soil organic carbon stock $dSOC_dt$ and measured changes of floatable soil fractions $dFloat_dt$ or cattle density $CattD$.

The Random Forest, which is a commonly used first machine learning model [46], was not significantly better (Figure S2). This indicates that it will not be possible to obtain a highly improved interpretable model with the given set of variables.

3.2. Soil Organic Carbon Equilibrium Level

According to the estimated version of Equation (4) given by Equation (5), we assess the actual ($p = 1$) steady-state soil organic carbon stock in the case where there are no changes in soil texture ($d\text{Float_dt} = 0$):

$$SOC_{eq} = \frac{0.189 + 0.009 TWI + 0.083 CattD^2}{0.004 + 0.002 AI - 0.002 CattD^2} \quad (6)$$

For mean values of explanatory variables, SOC_{eq} is equal to 51 t ha^{-1} , which means that growth of SOC sequestration (from the actual mean of 43 t ha^{-1}) can be expected in the long term if all variables remain constant.

4. Discussion

Most publications on the dynamics of changes in organic matter concern single locations, whereas there are fewer studies analyzing the dynamics of SOC changes over larger areas. A review of studies on changes in soil organic matter resources for the Netherlands [47] indicates a significant decrease in the SOC content in the 30–100 cm layer, but the study did not analyze changes in the function of factors that may affect them. Studies for England and Wales [36,48] show a decrease in SOC in soils with initially higher SOC content. Similar observations were made in Poland [32,49]. Analyses of changes in SOC in the UK show the usefulness of normalizing the SOC to the SOC to clay ratio and indicate that a realistic SOC/clay value for arable land is 1/13, and for grassland, it is greater than 1/8, furthermore showing that the largest decreases in the SOC/clay ratio occurred in soils with initially high levels [50]. Studies conducted in China [51] have shown the impact of the interaction of fertilization and the groundwater table on the changes in SOM.

Despite the existence of few studies analyzing changes in the level of soil organic matter in large areas, there is a clear deficit of analyses in which dynamic models have been developed that allow simulations of changes in SOC stocks in a longer timeframe. Our analysis partly fills this gap by offering an approach that combines mechanistic models based on differential equations with an empirical approach based on the analysis of changes using the estimation of simple linear relationships by regression methods. The proposed model provides a new method of extending knowledge on the mechanisms determining the amount of carbon supplied to the soil by interpreting the estimated equation in terms of accumulation and decomposition processes. This is possible due to the fact that the functional form of these processes is not defined a priori in the model, but is determined in the machine learning process [52]. It should be emphasized that the estimated model with the imposed structure of the first-order kinetics equation makes sense only when the terms I and k in the estimated version of formula 2 are positive in the model's application area. Otherwise, a model is obtained in which the simulated SOC level, despite a long period and fixed values of variables, does not tend to steady-state but, for example, disappears to zero or increases indefinitely. Such a result should most likely be interpreted as the effect of systematic errors in the measurements of SOC changes or layers of explanatory variables. Other mechanistic models do not offer the possibility to modify the variables describing processes of accumulation and decomposition, and some of their parameters require measurements on a local scale. Attempts to determine the parameters of the models used so far for studies on a macro scale led to significant errors in the estimation of crop carbon inputs [26]. Due to the large number of parameters, these models also pose a risk of overfitting [7].

The accuracy of the model estimated in this work conducted for agricultural soils in Poland ($R^2 = 0.05$; $RMSE = 0.5 \text{ t ha}^{-1} \text{ year}^{-1}$) is comparable to the accuracy of few other models for which annual changes of SOC ($\text{t ha}^{-1} \text{ year}^{-1}$) are known at point locations. In the case of the Yasso07 model used to predict changes in SOC stocks in 101 forest locations in France [53], the coefficient of determination was below 0.1. For the N14CP model adapted to agricultural areas, the test of predicting SOC changes on data from 11 long-term experiments conducted in Europe [54] showed a positive but insignificant relationship

($R^2 < 0.01$, $p > 0.05$). Numerous attempts have also been made to simulate changes in SOC stocks using machine learning methods such as Quantile Regression Forest (QRF), Random Forest (RF), or Gradient Boosting Model (GBM). These methods were used for areas in such countries as Hungary [55,56], Argentina [57], and China [58]. They were also used for selected arable fields in the central part of Nova Scotia in Canada [59]. Unfortunately, cross-validation prediction errors were most often reported for the predicted final SOC level (t ha^{-1}), and predicted SOC change rate errors were determined only for the model of SOC change in Hungary where $\text{RMSE} = 1.01 \text{ t ha}^{-1} \text{ year}^{-1}$ [56]. In these models, soil type maps were the most important covariates for Hungary, climatic variables for Argentina and China, and Landsat-based Tasseled Cap greenness for pre-growing periods for regional studies in Canada.

The obtained functional form of the model defined by Equation (5) and analysis of the signs of the estimated parameters of the equation, revealed that:

- the rate of change of SOC is proportional to the rate of change of the fraction of floatable particles, indicating a substantial shift of the sampling point from the previous location and/or soil leaching,
- an increase in cattle stocking density CattD increases the accumulation of post-harvest residues and reduces mineralization,
- an increase in the topographic-dependent part of soil moisture given by TWI increases the accumulation of post-harvest residues, while the increase in the climate-dependent part of soil moisture AI increases mineralization,
- the coefficients at p indicate that in older periods, the SOC accumulation has a lower rate but the change rate of SOC associated with the change rate of the fraction of floatable particles was higher,
- the relation between rates of change in SOC and floatable fractions is weaker for new changes, which may support a hypothesis on a problem with the sampling point localization in previous periods. This problem seemed to be less relevant as the localization equipment (GPS) developed.

The value of the regression coefficient for the rate of change of soil floatable fraction dFloat_dt for older periods can probably be mainly associated with errors in point localization because the use of GPS was only possible for measurements in newer periods: 2002–2004 and 2018.

The secondary model for the equilibrium SOC level, given by Equation (6), predicts high SOC content in river valleys and cattle farming areas, which is in line with our expectations. In grassland areas in Poland, a positive correlation between cattle stocking density and the percentage of organic carbon in the soil is known [60]. However, the model also predicts that as the Aridity Index (AI) increases, the equilibrium SOC content decreases, which contradicts most observations that SOC levels are higher in wet climates than in dry ones. In addition, expected variables such as the share of floatable fraction (Float) or the sum of degree-days (GDD5) were not included in the model, although there is an observed and physically justified relationship between the content of clay fractions and the level of SOC [10] and the relationship between temperature and the rate of SOC decomposition [4]. A similar problem concerns the lack of the LS variable in the model, which describes the topographic component of water erosion and should be present in the model due to erosion being one of the main mechanisms of carbon loss from soils [61]. The values of coefficients that are inconsistent with expectations may be related to too few measurement points and their limited spatial range. This means that some variables, locally correlated with other variables not taken into account in the model but important for the modeled process (omission error), become their impossible-to-interpret substitute.

Problems with proper recognition of the climate impact on the model indicate that the model is estimated on data from four regions of Poland and show that it is restricted to the study area. In order to obtain a more general model in which the influence of climatic factors on the SOC would be correctly recognized, calibration across several climatic zones and on the longer time series of measurements is needed.

The proposed model can be used wherever SOC stocks have been measured for several dozen point locations in at least two distant moments in time. The only limitation for the use of the model in other conditions, e.g., in other parts of the world, may be significant spatial aggregation of the available variables explaining the use of natural fertilizers or the amount of post-harvest residues. The explanatory variables selected in this study in Poland can be replaced with similarly interpreted variables available in the study region. The presented model for Poland is, therefore, rather an example of the implementation of a new method of creating SOC dynamics models than a typical ready-to-use simulation model calibrated for all conditions.

5. Conclusions

A model with the features of a mechanistic model was developed to enable a simple estimation of its parameters using the regularized LASSO regression method. The explicit form of the model is in the form of a differential equation, which allows for the simulation of changes in SOC resources in the long term. It also enables the determination of the equilibrium SOC level, the analysis of which for the examined points indicated that there is potential for carbon sequestration in the soils across the studied regions of Poland.

Analysis of the equation for SOC change rate and the long-term equilibrium SOC level showed that for selected regions in Poland:

- the stocking density of cattle increases the accumulation of post-harvest residues and reduces mineralization; in older analyzed periods, accumulation has a lower rate;
- old measurements of SOC are affected by problems with exact sampling point localization, which can be reflected in models by a term with a level of change of soil floatable fraction which, in principle, should not change if the soil profile is not physically transformed;
- the equilibrium SOC level predicted from our model is high in areas with high TWI e.g., river valleys or areas with high cattle density, which is in line with expectations.

The weakness of the estimated model for Poland lies in the lack of inclusion of important variables, such as soil texture, whose impact on SOC is known from previous studies, or the inclusion of some variables, such as AI, with a sign inconsistent with expectations. This problem is probably related to the small spatial spread of the sampling area, which indicates the need to create and use databases containing measurements on diverse soils located in various climatic zones.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/land12081587/s1>, Figure S1: Diagnostic plots of LASSO regression; Figure S2: Permutation importance plot and partial dependence plots of RF model; Table S1: Regression LASSO model results vs OLS regression.

Author Contributions: Conceptualization, A.Ł., G.S., T.S. and R.K.; methodology, A.Ł.; formal analysis, A.Ł.; investigation, R.K. and A.Ł.; writing—original draft preparation, A.Ł.; writing—review and editing, A.Ł., G.S., T.S. and R.K.; visualization, A.Ł.; supervision, G.S. and T.S.; project administration, G.S.; funding acquisition, G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre, Poland, the project number 2016/21/N/ST10/02649.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author under reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zomer, R.J.; Bossio, D.A.; Sommer, R.; Verchot, L.V. Global Sequestration Potential of Increased Organic Carbon in Cropland Soils. *Sci. Rep.* **2017**, *7*, 15554. [CrossRef] [PubMed]
2. Xu, S.; Sheng, C.; Tian, C. Changing soil carbon: Influencing factors, sequestration strategy and research direction. *Carbon Balance Manag.* **2020**, *15*, 2. [CrossRef] [PubMed]
3. Powlson, D.S.; Poulton, P.R.; Glendining, M.J.; Macdonald, A.J.; Goulding, K.W.T. Is it possible to attain the same soil organic matter content in arable agricultural soils as under natural vegetation? *Outlook Agric.* **2022**, *51*, 91–104. [CrossRef]
4. De Willigen, P.; Janssen, B.H.; Heesmans, H.I.M.; Conijn, J.G.; Velthof, G.L.; Chardon, W.J. Decomposition and Accumulation of Organic Matter in Soil; Comparison of Some Models (No.1726), Alterra. 2008. ISSN 1566-7197. Available online: <https://edepot.wur.nl/15401> (accessed on 15 May 2023).
5. McGill, W.B. Review and classification of ten Soil Organic Matter (SOM) models. In *Evaluation of Soil Organic Matter Models*; Powlson, D.S., Smith, P., Smith, J.U., Eds.; NATO ASI Series; Springer: Berlin/Heidelberg, Germany, 1996; Volume 38, pp. 111–132. [CrossRef]
6. Manzoni, S.; Porporato, A. Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biol. Biochem.* **2009**, *41*, 1355–1379. [CrossRef]
7. Le Noë, J.; Manzoni, S.; Abramoff, R.; Bölscher, T.; Bruni, E.; Cardinael, R.; Ciais, P.; Chenu, C.; Clivot, H.; Derrien, D.; et al. Soil organic carbon models need independent time-series validation for reliable prediction. *Commun. Earth Environ.* **2023**, *4*, 158. [CrossRef]
8. Todd-Brown, K.E.O.; Hopkins, F.M.; Kivlin, S.N.; Talbot, J.M.; Allison, S.D. A framework for representing microbial decomposition in coupled climate models. *Biogeochemistry* **2012**, *109*, 19–33. [CrossRef]
9. Georgiou, K.; Abramoff, R.Z.; Harte, J.; Riley, W.J.; Torn, M.S. Microbial community-level regulation explains soil carbon responses to long-term litter manipulations. *Nat. Commun.* **2017**, *8*, 1223. [CrossRef] [PubMed]
10. White, C.M.; Kemanian, A.R.; Kaye, J.P. Implications of carbon saturation model structures for simulated nitrogen mineralization dynamics. *Biogeosciences* **2014**, *11*, 6725–6738. [CrossRef]
11. Xu, X.; Shi, Z.; Li, D.; Rey, A.; Ruan, H.; Craine, J.M.; Liang, J.; Zhou, J.; Luo, Y. Soil properties control decomposition of soil organic carbon: Results from data-assimilation analysis. *Geoderma* **2016**, *262*, 235–242. [CrossRef]
12. Luo, Y.; Ahlström, A.; Allison, S.D.; Batjes, N.H.; Brovkin, V.; Carvalhais, N.; Chappell, A.; Ciais, P.; Davidson, E.A.; Finzi, A.; et al. Toward more realistic projections of soil carbon dynamics by Earth system models, Global Biogeochem. Cycles **2016**, *30*, 40–56. [CrossRef]
13. Lawrence, C.R.; Neff, J.C.; Schimel, J.P. Does adding microbial mechanisms of decomposition improve soil organic matter models? A comparison of four models using data from a pulsed rewetting experiment. *Soil Biol. Biochem.* **2009**, *41*, 1923–1934. [CrossRef]
14. Wieder, W.R.; Allison, S.D.; Davidson, E.A.; Georgiou, K.; Hararuk, O.; He, Y.; Hopkins, F.; Luo, Y.; Smith, M.J.; Sulman, B.; et al. Explicitly representing soil microbial processes in earth system models. *Glob. Biogeochem. Cycles* **2015**, *29*, 1782–1800. [CrossRef]
15. Shi, Z.; Crowell, S.; Luo, Y.; Moore, B. Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nat. Commun.* **2018**, *9*, 2171. [CrossRef] [PubMed]
16. Menichetti, L.; Ågren, G.I.; Barré, P.; Moyano, F.; Kätterer, T. Generic parameters of first-order kinetics accurately describe soil organic matter decay in bare fallow soils over a wide edaphic and climatic range. *Sci. Rep.* **2019**, *9*, 20319. [CrossRef] [PubMed]
17. Campbell, E.E.; Paustian, K. Current developments in soil organic matter modeling and the expansion of model applications: A review. *Environ. Res. Lett.* **2015**, *10*, 123004. [CrossRef]
18. Coleman, K.; Jenkinson, D.S. RothC-26.3—A Model for the turnover of carbon in soil. In *Evaluation of Soil Organic Matter Models*; Powlson, D.S., Smith, P., Smith, J.U., Eds.; NATO ASI Series; Springer: Berlin/Heidelberg, Germany, 1996; Volume 38, pp. 237–246. [CrossRef]
19. Falloon, P.; Smith, P. Simulating SOC changes in long-term experiments with RothC and CENTURY: Model evaluation for a regional scale application. *Soil Use Manag.* **2002**, *18*, 101–111. [CrossRef]
20. Morais, T.G.; Teixeira, R.F.; Domingos, T. Detailed global modelling of soil organic carbon in cropland, grassland and forest soils. *PLoS ONE* **2019**, *14*, e0222604. [CrossRef]
21. Parton, W.J.; Schimel, D.S.; Cole, C.V.; Ojima, D.S. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Sci. Soc. Am. J.* **1987**, *51*, 1173–1179. [CrossRef]
22. Dimassi, B.; Guenet, B.; Saby, N.P.; Munoz, F.; Bardy, M.; Millet, F.; Martin, M.P. The impacts of CENTURY model initialization scenarios on soil organic carbon dynamics simulation in French long-term experiments. *Geoderma* **2018**, *311*, 25–36. [CrossRef]
23. Li, C.; Frolking, S.; Frolking, T.A. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. *J. Geophys. Res. Atmos.* **1992**, *97*, 9759–9776. [CrossRef]
24. Lembaid, I.; Moussadek, R.; Mrabet, R.; Bouhaouss, A. Modeling Soil Organic Carbon Changes under Alternative Climatic Scenarios and Soil Properties Using DNDC Model at a Semi-Arid Mediterranean Environment. *Climate* **2022**, *10*, 23. [CrossRef]
25. Robinson, T.P.; Wint, G.R.W.; Conchedda, G.; Van Boeckel, T.P.; Ercoli, V.; Palamara, E.; Cinardi, G.; D’Aietti, L.; Hay, S.; Gilbert, M. Mapping the Global Distribution of Livestock. *PLoS ONE* **2014**, *9*, e96084. [CrossRef] [PubMed]
26. Keel, S.G.; Leifeld, J.; Mayer, J.; Taghizadeh-Toosi, A.; Olesen, J.E. Large uncertainty in soil carbon modelling related to method of calculation of plant carbon input in agricultural systems. *Eur. J. Soil Sci.* **2017**, *68*, 953–963. [CrossRef]

27. Todd-Brown, K.E.O.; Randerson, J.T.; Post, W.M.; Hoffman, F.M.; Tarnocai, C.; Schuur, E.A.G.; Allison, S.D. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **2013**, *10*, 1717–1736. [\[CrossRef\]](#)
28. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **2018**, *5*, 180214. [\[CrossRef\]](#) [\[PubMed\]](#)
29. D'andrimont, R.; Verhegghen, A.; Lemoine, G.; Kempeneers, P.; Meroni, M.; van der Velde, M. From parcel to continental scale—A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. *Remote Sens. Environ.* **2021**, *266*, 112708. [\[CrossRef\]](#)
30. Shamrikova, E.; Kondratenok, B.; Tumanova, E.; Vanchikova, E.; Lapteva, E.; Zonova, T.; Lu-Lyan-Min, E.; Davydova, A.; Libohova, Z.; Suvannang, N. Transferability between soil organic matter measurement methods for database harmonization. *Geoderma* **2022**, *412*, 115547. [\[CrossRef\]](#)
31. Warzyński, H.; Sosnowska, A.; Harasimiuk, A. Effect of variable content of organic matter and carbonates on results of determination of granulometric composition by means of Casagrande's areometric method in modification by Prószyński. *Soil Sci. Annu.* **2018**, *69*, 39–48. [\[CrossRef\]](#)
32. Kaczynski, R.; Siebielec, G.; Hanegraaf, M.C.; Korevaar, H. Modelling soil carbon trends for agriculture development scenarios at regional level. *Geoderma* **2017**, *286*, 104–115. [\[CrossRef\]](#)
33. Krasuska, E.; Faber, A.; Jarosz, Z.; Kaczyński, R. Modelling soil organic carbon sequestration under crops for biofuels in Poland. *Pol. J. Agron.* **2016**, *27*, 126–136. [\[CrossRef\]](#)
34. Kopiński, J.; Witorożec-Piechnik, A. Assessment of Soil Organic Matter Management in Polish Agriculture. *Annals PAAAE* **2022**, *XXIV*, 44–54. [\[CrossRef\]](#)
35. Jenkinson, D.S. The turnover of organic carbon and nitrogen in soil. *Philos. Trans. R. Soc. Lond. B* **1990**, *329*, 361–388. [\[CrossRef\]](#)
36. Kirk, G.J.D.; Bellamy, P.H. Analysis of changes in organic carbon in mineral soils across England and Wales using a simple single-pool model. *Eur. J. Soil Sci.* **2010**, *61*, 406–411. [\[CrossRef\]](#)
37. Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932–3937. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [\[CrossRef\]](#)
39. EU-DEM, v.1.1, Copernicus. Available online: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata> (accessed on 4 December 2022).
40. Kopecký, M.; Macek, M.; Wild, J. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Sci. Total Environ.* **2021**, *757*, 143785. [\[CrossRef\]](#) [\[PubMed\]](#)
41. IMiGW. Meteorological Data. Available online: https://danepubliczne.imgw.pl/data/dane_pomiarowo_obserwacyjne/dane_meteorologiczne/ (accessed on 4 December 2022).
42. PSR 2002. Bank Danych Lokalnych, Powszechne Spisy Rolne, PSR2002. Available online: <https://bd.stat.gov.pl/bdl/dane/podgrup/tematselekcja:AGRICULTURALCENSUS2002-BYHOLDINGSHEADQUARTER/Agriculturalholdingsbytypeandlivestockpopulation/2002/agriculture,grandtotal/cattle> (accessed on 10 April 2023).
43. PSR 2020. Bank Danych Lokalnych, Powszechne Spisy Rolne, PSR2020. Available online: <https://bd.stat.gov.pl/bdl/dane/podgrup/tematselekcja:AGRICULTURALCENSUS2020-BYHOLDINGSHEADQUARTER/Livestock/2020/agriculturalholdingstotal/animalpopulation/cattletotal> (accessed on 10 April 2023).
44. Siebielec, G. Stały monitoring gleb użytków rolnych Polski. *Stud. Rap. IUNG-PIB* **2017**, *51*, 57–72. [\[CrossRef\]](#)
45. GIOŚ. Monitoring Chemizmu Gleb Ornych Polski. Available online: https://www.gios.gov.pl/chemizm_gleb/ (accessed on 18 January 2023).
46. Tyralis, H.; Papacharalampous, G.; Langousis, A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **2019**, *11*, 910. [\[CrossRef\]](#)
47. Knotters, M.; Teuling, K.; Reijneveld, A.; Lesschen, J.P.; Kuikman, P. Changes in organic matter contents and carbon stocks in Dutch soils, 1998–2018. *Geoderma* **2022**, *414*, 115751. [\[CrossRef\]](#)
48. Bellamy, P.H.; Loveland, P.J.; Bradley, R.I.; Lark, R.M.; Kirk, G.J.D. Carbon losses from all soils across England and Wales 1978–2003. *Nature* **2005**, *437*, 245–248. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Siebielec, G.; Łopatka, A.; Smreczak, B.; Kaczyński, R.; Siebielec, S.; Koza, P.; Dach, J. Materia organiczna w glebach mineralnych Polski. *Stud. Rap. IUNG-PIB* **2020**, *64*, 9–30. [\[CrossRef\]](#)
50. Prout, J.M.; Shepherd, K.D.; McGrath, S.P.; Kirk, G.J.D.; Hassall, K.L.; Haefele, S.M. Changes in organic carbon to clay ratios in different soils and land uses in England and Wales over time. *Sci. Rep.* **2022**, *12*, 5162. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Wu, Z.; Liu, Y.; Li, G.; Han, Y.; Li, X.; Chen, Y. Influences of Environmental Variables and Their Interactions on Chinese Farmland Soil Organic Carbon Density and Its Dynamics. *Land* **2022**, *11*, 208. [\[CrossRef\]](#)
52. Grunwald, S. Artificial intelligence and soil carbon modeling demystified: Power, potentials, and perils. *Carbon Footpr.* **2022**, *1*, 5. [\[CrossRef\]](#)
53. Mao, Z.; Derrien, D.; Didion, M.; Liski, J.; Eglin, T.; Nicolas, M.; Jonard, M.; Saint-André, L. Modeling soil organic carbon dynamics in temperate forests with Yasso07. *Biogeosciences* **2019**, *16*, 1955–1973. [\[CrossRef\]](#)
54. Janes-Bassett, V.; Davies, J.; Rowe, E.C.; Tipping, E. Simulating long-term carbon nitrogen and phosphorus biogeochemical cycling in agricultural environments. *Sci. Total Environ.* **2020**, *714*, 136599. [\[CrossRef\]](#)

55. Szatmári, G.; Pirkó, B.; Koós, S.; Laborczi, A.; Bakacsi, Z.; Szabó, J.; Pásztor, L. Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. *Soil Tillage Res.* **2019**, *195*, 104410. [[CrossRef](#)]
56. Szatmári, G.; Pásztor, L.; Heuvelink, G.B. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. *Geoderma* **2021**, *403*, 115356. [[CrossRef](#)]
57. Heuvelink, G.B.M.; Angelini, M.E.; Poggio, L.; Bai, Z.; Batjes, N.H.; Van den Bosch, R.; Bossio, D.; Estella, S.; Lehmann, J.; Olmedo, G.F.; et al. Machine learning in space and time for modelling soil organic carbon change. *Eur. J. Soil. Sci.* **2021**, *72*, 1607–1623. [[CrossRef](#)]
58. Li, H.; Wu, Y.; Liu, S.; Xiao, J.; Zhao, W.; Chen, J.; Alexandrov, G.; Cao, Y. Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Glob. Change Biol.* **2022**, *28*, 3394–3410. [[CrossRef](#)] [[PubMed](#)]
59. Paul, S.S.; Heung, B.; Lynch, D.H. Modeling of total and active organic carbon dynamics in agricultural soil using digital soil mapping: A case study from Central Nova Scotia. *Can. J. Soil Sci.* **2022**, *103*, 64–80. [[CrossRef](#)]
60. Pietrzak, S.; Hołaj-Krzak, J.T. The content and stock of organic carbon in the soils of grasslands in Poland and the possibility of increasing its sequestration. *J. Water Land Dev.* **2022**, *54*, 68–76. [[CrossRef](#)]
61. Naipal, V.; Ciais, P.; Wang, Y.; Lauerwald, R.; Guenet, B.; Van Oost, K. Global soil organic carbon removal by water erosion under climate change and land use change during AD 1850–2005. *Biogeosciences* **2018**, *15*, 4459–4480. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.